

## Claims

We claim:

1. A method for identifying a location of an object in a physical scene with a stereo camera comprising:

identifying a virtual surface in the physical scene;  
constructing an approximate disparity set for the virtual surface;  
acquiring a main and a reference image of the scene with the stereo camera;  
warping the reference image according to the disparity set;  
subtracting pixels of the warped reference image from corresponding pixels

of the main image to determine a depth residual of each pixel; and

identifying pixels having a substantially non-zero residual with a surface of the object not coincident with the virtual surface.

2. The method of claim 1 wherein the virtual surface has an associated margin to form a virtual volume near the virtual surface with a thickness equal to the margin.

3. The method of claim 1 wherein the virtual surface is an arbitrary surface defined in a space of the physical scene.

4. The method of claim 1 wherein the virtual surface is partially tangible and partially in a space of the scene.

5. The method of claim 1 further comprising:

setting each depth residual less than a predetermined threshold to zero; and

setting all other depth residuals to one to generate a binary segmentation mask for the object.

6. The method of claim 1 wherein the object is moving, and further comprising:  
tracking the moving object in a stereo video of the scene using the binary segmentation mask.

7. The method of claim 1 further comprising:  
acquiring a sparse set of point correspondences from a calibration pair of images;  
applying a polynomial interpolation to the sparse set of point correspondences to generate a smooth continuous approximate disparity set.

8. The method of claim 7 wherein a particular disparity,  $\mathbf{d}(x, y)$  is approximated by a linear system  $\mathbf{d}(x, y) = \Lambda \tilde{\mathbf{x}}(x, y)$ , where  $\Lambda$  is an unknown matrix of coefficients, and  $\tilde{\mathbf{x}}(x, y)$  is a power expansion of  $\mathbf{x} = [x, y]^T$

$$\tilde{\mathbf{x}}(x, y) = \begin{bmatrix} x^2 \\ y^2 \\ xy \\ x \\ y \\ 1 \end{bmatrix}.$$

9. The method of claim 1 wherein the virtual surface is substantially planar and the approximated disparity set is obtained from intrinsic camera parameters of the stereo camera.

10. The method of claim 1 further comprising:

determining a touching of the virtual surface by the object from the depth disparities.

11. The method of claim 1 further comprising:

illuminating the scene and the object with a dynamic projector.

12. The method of claim 11 wherein the illumination includes a high contrast image.

13. The method of claim 2 further comprising:

performing volumetric depth segmentation operations according to virtual volume.

14. The method of claim 1 further comprising:

identifying a first virtual surface in the physical scene;  
identifying a second virtual surface in the physical scene offset from the first virtual surface by a constant distance;

analytically constructing an approximate disparity set for the first virtual surface and the second virtual surface;

warping the reference image according to the first disparity set;  
warping the reference image according to the second disparity set;  
subtracting each pixel of the first warped reference image from a corresponding pixel of the main image to determine a first depth residual of each pixel; and

subtracting each pixel of the second reference image from a corresponding pixel of the main image to determine a second depth residual of each pixel; and

comparing the first and second depth residuals to determine a touching of the virtual surface.

15. A system for identifying a location of an object in a physical scene comprising:

means for identifying a virtual surface in the physical scene;

means for analytically constructing an approximate disparity set for the virtual surface;

a stereo camera acquiring a main and a reference image of the scene;

means for warping the reference image according to the disparity set;

means for subtracting each pixel of the warped reference image from a corresponding pixel of the main image to determine a depth residual of each pixel; and

means for identifying each pixel having a substantial non-zero depth residual with a surface of the object not coincident with the virtual surface.

16. The system of claim 15 wherein the virtual surface has an associated margin to form a virtual volume near the virtual surface with a thickness equal to the margin.

17. The system of claim 15 wherein the virtual surface is an arbitrary surface defined in a space of the physical scene.

18. The system of claim 15 wherein the virtual surface is partially tangible and partially in a space of the scene.

19. The system of claim 15 further comprising:

means for setting each depth residual less than a predetermined threshold to zero; and

means for setting all other depth residuals to one to generate a binary segmentation mask for the object.

20. The system of claim 15 wherein the object is moving, and further comprising:  
means for tracking the moving object in a stereo video of the scene using the binary segmentation mask.

20190111